A gold mine, but still no Klondike

Nordic register data in health inequalities research

Van Der Wel, Kjetil A.; Östergren, Olof; Lundberg, Olle; Korhonen, Kaarina; Martikainen, Pekka; Andersen, Anne-Marie Nybo; Urhøj, Stine Kjær

Published in:
Scandinavian Journal of Public Health

DOI:
10.1177/1403494819858046

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Scandinavian Journal of Public Health, 1–13

ORIGINAL ARTICLE

A gold mine, but still no Klondike: Nordic register data in health inequalities research

KJETIL A. VAN DER WEL[1], OLOF ÖSTERGREN[2], OLLE LUNDBERG[2], KAARINA KORHONEN[3], PEKKA MARTIKAINEN[3], ANNE-MARIE NYBO ANDERSEN[4] & STINE KJAER URHOJ[4]

[1]*Department of Social Work, Child Welfare and Social Policy, OsloMet – Oslo Metropolitan University, Norway,* [2]*Department of Public Health Sciences, Stockholm University, Sweden,* [3]*Faculty of Social Sciences, Population Research Unit, University of Helsinki, Finland,* [4]*Department of Public Health, Section of Epidemiology, University of Copenhagen, Denmark*

Abstract

Aims: Future research on health inequality relies on data that cover life-course exposure, different birth cohorts and variation in policy contexts. Nordic register data have long been celebrated as a 'gold mine' for research, and fulfil many of these criteria. However, access to and use of such data are hampered by a number of hurdles and bottlenecks. We present and discuss the experiences of an ongoing Nordic consortium from the process of acquiring register data on socio-economic conditions and health in Denmark, Finland, Norway and Sweden. *Methods*: We compare experiences of data-acquisition processes from a researcher's perspective in the four countries and discuss the comparability of register data and the modes of collaboration available to researchers, given the prevailing ethical and legal restrictions. *Results*: The application processes we experienced were time-consuming, and decision structures were often fragmented. We found substantial variation between the countries in terms of processing times, costs and the administrative burden of the researcher. Concerned agencies differed in policy and practice which influenced both how and when data were delivered. These discrepancies present a challenge to comparative research. ***Conclusions*: We conclude that there are few signs of harmonisation, as called for by previous policy documents and research papers. Ethical vetting needs to be centralised both within and between countries in order to improve data access. Institutional factors that seem to facilitate access to register data at the national level include single storage environments for health and social data, simplified ethical vetting and user guidance.**

Keywords: *Register data, health inequality, Nordic countries, data, methodology*

Introduction

Research on public health and socio-economic inequalities in health strongly suggests that a range of economic and social factors contribute to the distribution of mortality and morbidity [1,2]. This has also been firmly established by recent appointed commissions in the Nordic countries [3–5]. The literature shows that the chances of staying healthy is clearly linked to a range of socially patterned exposures and resources accumulated over time, including childhood disadvantage, labour-market participation, working conditions and incomes and economic resources. Cohort-specific experiences and exposures during critical life stages [6] might also have a different impact on living conditions and health in different social strata. Furthermore, individuals' health and socio-economic position are interrelated over time; health affects socio-economic achieve-

Correspondence: Kjetil A. van der Wel, Department of Social Work, Child Welfare and Social Policy, OsloMet – Oslo Metropolitan University, Postboks 4, St. Olavs plass, Oslo, 0130, Norway. E-mail: Kjetil.wel@oslomet.no

Date received 13 March 2019; reviewed 24 May 2019; accepted 27 May 2019

DOI: 10.1177/1403494819858046
journals.sagepub.com/home/sjp

ment, which in turn shapes health and future social-mobility opportunities.

The insight that a broad range of conditions are involved in shaping inequalities in health has led attention to welfare-state arrangements more generally and to specific policies such as unemployment benefits [7]. The logic behind this is that the design of welfare policies, including education, social protection and labour-market policies, has the potential to affect inequalities in conditions and opportunities and thereby also health inequalities. Yet, we need more knowledge about how health inequalities are shaped across life courses [8], and to what extent these processes can actually be modified by welfare policies [9–11]. In order to address these complex relationships, there is a need for theoretical and methodological development [8], which in turn drives the need for high-quality data.

The Nordic countries form an interesting case here, partly because they have been assumed to have smaller inequalities (the so-called Nordic paradox) [12–14], but mainly because of a basic similarity between the countries combined with considerable variation in the design of welfare policies and rich sets of administrative register data covering these policies as well as some key conditions of life. These registers contain detailed and reliable information on people's economic and social conditions as well as their health, and cover long periods of the life course for a large number of people from different birth cohorts. This means that a number of the problems that characterise survey data often used [15], such as non-response and small samples restricting the possibilities to study smaller groups, can be avoided. Using Nordic register data in a comparative approach will therefore allow us to address key issues better regarding the interplay between welfare states and health through analyses of how people's social circumstances and opportunities are interlinked with health over time, and how different policy options are more or less able to break, or at least ameliorate, these relationships.

While Nordic register data have been used extensively for national and to some extent comparative research already, there is still significant unexploited potential. This has also been identified as a strategic issue for research policy in the Nordic region, and Nordic register data have been praised as a 'gold mine' for research [16–18]. However, despite its celebrated position in research policy documents, register data on health and socio-economic conditions are not easily available to national or international research groups, and they are not immediately ready to use or instantly comparable, as they are restricted to what can be seen through the lens of the welfare state. Thus, rather than picking gold nuggets laying bare, researchers may find themselves digging deep for gold in harsh conditions. The objective of the current paper is to provide practical insights to policy-makers, register owners and researchers about the main barriers that impede comparative research using the Nordic register data.

What are Nordic register data?

A register consists of data that are collected as part of welfare-state 'book-keeping', from basic information about people's addresses and family units, their education and taxable incomes to use of social insurance schemes or health care. Such registers are kept by different agencies, but a key feature in the Nordic countries is the use of a unique personal identification number (PIN) used in all public registers, as well as by employers, banks and most other private companies. The PIN therefore makes it possible to link information from different sources regarding each individual, but through meta-registers it is also possible to link individuals to their parents, children and spouses and to place of residence [19]. Register data are typically collected for administrative purposes, and the data content will thus reflect administrative needs. Like most survey data, register data do not include information on undocumented immigrants.

The Nordic countries have a long history of collecting register data regarding births, deaths, diseases, emigrations and immigrations, and various social issues for administrative purposes [19]. All of the Nordic countries have a Central Person Register (CPR) with the PIN for all residents and information on date of birth, sex, marital status and so on [20]. Additionally, each country has a set of other registers covering social, family and labour-market characteristics, for example educational registers, taxation registers, income registers, labour-market registers, health registers, birth registers and causes-of-death registers [20].

Although these registers are designed and used for administrative purposes, they have also become an important data source for research. Register data include extensive information on the total populations in the Nordic countries – about 26 million individuals in total [21,22]. This information is collected in a uniform manner for all individuals across their life courses. Since the registration of information is in general not dependent on consent of participation [21], attrition and non-response is not an issue. These features make register data well suited to the needs of future health-inequality research. Like all secondary data, however, they also have limitations,

including the lack of theoretically based variables, no subjective information on health and attitudes, changing reporting practices over time and reporting errors.

Nordic research policy

Nordic register data provide ample opportunities for a range of research questions and study designs. In acknowledgement of this, a number of initiatives have been launched in order to expose and overcome some of the challenges that impede collaborative efforts using register data. The Nordic Council and its research council NordForsk have tried to promote Nordic integration in the area of register-based research. For instance, since 2006, there has been ongoing work with e-science, which includes the ambition to facilitate increased use of Nordic registers in research [23]. The Könberg report to the Nordic Council [16] proposed co-ordinated ethical vetting, easier data sharing across borders and constructing Nordic research registers. The report also outlines an idea for a 'virtual data centre' for register research that can facilitate research and collaboration. Other initiatives include policy groups, networks, pilot projects and surveys [23]. In particular, a Nordic platform for sensitive data (Tryggve) and a Nordic Microdata Access Network (NordMAN) has been established. National initiatives by research councils have also emphasised the importance of strengthening register-based research and the benefits of Nordic collaboration, harmonisation and coordination. It remains, however, to be seen how these initiatives will impact the conditions of comparative register-based research once implemented. This paper contributes with up-to-date insights on the accessibility of register data for research.

Aims

We aim to identify and discuss various kinds of obstacles to register-based research from a user's point of view. We draw on real-life experiences from an ongoing Nordic consortium in order to contrast the promise of the gold mine with the actual experiences of mining. More specifically, we aim to describe and contrast the data-acquisition process across Denmark, Finland, Norway and Sweden, focusing on bottlenecks and hurdles, and discuss how these differ between countries. We will also highlight some types of comparability problems related to the data itself, as well as some practical modes of conducting comparative analyses, given constraints. Finally, we aim to discuss how the various bottlenecks and hurdles that we have experienced and documented may impede research, and to what extent there may be ways to make improvements.

Analytical approach

The basis for this work was a NordForsk call for Register Pilots in 2014 that took an empirical approach to the problems that researchers who want to use registers for comparative purposes may face. One specific aim of this call was '…to support projects targeting the construction of Nordic research infrastructures in the health and welfare sphere by combining socio-economic and health-related registers', while another aim of the call was to '…monitor existing hurdles and bottlenecks (technical, organisational, legal, ethical, etc.) that impede Nordic research and research infrastructure cooperation'.

Our consortium of researchers from Norway, Denmark, Finland and Sweden saw this call as an opportunity to access new register data for comparative work while at the same time documenting the process and problems occurring throughout this process. Our application for the project Contingent Lifecourses (C-LIFE; see https://blogg.hioa.no/clife) was funded by the NordForsk Register Pilot call (project number 75970).

The starting point differed between the members of the consortium. The Norwegian and Swedish teams had some access to older data but needed to go through the process from the start to acquire new and updated data. This was also the case for the Danish team, but the Danish team had some advantages in relation to the application process due to a special set-up at their institution. The Finnish team had access to two data sets that included register data on population-based samples but needed to update and extend these data. While these differences in preconditions make it more difficult to compare the experiences, they also provide some evidence on the structural differences in modes of data access that exist across the Nordic countries.

In order to generate the experiences necessary to document hurdles and bottlenecks, we pursued two strategies: (a) going through the process of acquiring register data for future research and documenting the steps, and (b) writing a series of scientific papers on the basis of existing register data available to the teams to explore different ways of doing comparative work within the legal and ethical frameworks that apply. The national data applications were not identical but were adapted to the broader research needs of each research environment. Pragmatic and strategic considerations also contributed to some differences regarding data sources (such as the inclusion of prescription medication data in Finland), the omission

of certain variables (such as geographical information in Norway) and the use of samples in Finland. All applications included a wide range of socio-economic and demographic variables on individuals and households, as well as hospitalisation data, cancer registries, medical birth registers and cause-of-death registers. While some of the limitations and differences are specific to the conditions of the C-LIFE project, the general type and nature of these are common to all researchers seeking to access register data. We thus argue that our experiences are valuable for generating some general insights into some of the issues that might impede register-based research using health and socio-economic data.

Obtaining data and making them comparable

This section first presents the data-acquisition process in each country. We then summarise and compare experiences before briefly discussing issues of comparability and different potential modes of collaboration.

Denmark

In Denmark, there are two main 'entries' to microdata from the Danish national registers: Statistics Denmark and the Danish Health Data Authority. Statistics Denmark has an extensive collection of register data on the total population, including information on various demographic factors and social conditions. The Danish Health Data Authority holds all health registers, including the National Patient Register and Causes of Death Registry. However, Statistics Denmark holds copies of important health registers, and when health data and data on social conditions are requested within the same project, Statistics Denmark extract data from the registers and place all the data on a server at Statistics Denmark.

An application for data available from Statistics Denmark (including copies of health registers) is sent only to Statistics Denmark, while applications for data from registers held by other institutions or authorities (including health data that are not available from Statistics Denmark) are sent separately to these institutions. The processing of these data applications includes ethical vetting. Authorisations from ethical committees are in general not required for register research. However, the processing and linkage of data have to be approved by the Danish Data Protection Agency. The legal basis for the processing of personal data is (from May 2018) the General Data Protection Regulation (GDPR).

The process of acquiring register data reported here deviates somewhat from other research environments in Denmark. The Department of Public Health at the University of Copenhagen has developed a common infrastructure on a separate server placed at Statistics Denmark, called the Public Health Database (Folkesundhedsdatabasen). This database comprises data from various registers at Statistics Denmark comprising information covering the entire Danish population since 1980. Researchers at the department can apply for access to data at the Public Health Database for specific projects. This usually makes processing time shorter than in other situations. Since the establishment of the Public Health Database, two projects have been established to facilitate Nordic collaboration on register research. The process of the establishment of a project is described below.

The first step of acquiring data was to prepare a research protocol, including a detailed description of the purpose of the study, the study population and a list of registers and specific variables. This was done in collaboration with data managers at the Public Health Database, employed at the Department of Public Health. The application forms and variable lists were processed by the researchers and data managers for approximately two months. This was done to define the study design, the population and period precisely and to decide the variables that were necessary following the 'need-to-know' principle. The final research protocol and variable list was sent to the Public Health Database steering committee, together with a research proposal to the Statistics Denmark. Both had to approve the research project.

While the applications were processed at Statistics Denmark and the Public Health Database, other approvals were sought and obtained. At the University of Copenhagen, the application to the Danish Data Protection Agency is sent to a 'Fællesanmeldelse' ('joint notification') at the University, which shortens the processing time, as the university approves the projects and register linkage. As part of this approval, the GDPR requires a Data Protection Impact Analysis (DPIA) when processing sensitive data (including information about health), vulnerable groups (including children) or large populations. The DPIA has to be approved by the university's Data Protection Officer (DPO). The DPIA was written by the researchers and approved by the DPO after four days. The application to the Danish Data Protection Agency (the 'joint notification') was approved after 16 days (including the four days where the DPIA was processed by the DPO). Statistics Denmark does not hold a copy of the Danish Cancer Registry, and the Danish Health Data Authority had to approve the

use of these data. The application was approved after 28 days.

When all other approvals were obtained, the Public Health Database steering committee and Statistics Denmark approved the project, and after approximately three months from submission of the application to the Public Health Database and Statistics Denmark, the data were extracted and placed on the remote-access platform ready to be linked and processed by the approved researchers. Statistics Denmark pseudonymised the national PINs and assigned an ID number in order to enable linkage of the data from the various registers. In addition, indirect identification of individuals is prevented by strict rules about outputs and publication of results.

For the Danish team, acquiring access to register data on health and social conditions involved obtaining approvals separately from several authorities, but the processing times were in general short. The planning phase and the preparation of the research proposal and variable list conducted together with the Public Health Database data managers was long compared to the rest of the process. However, this process was completed in order for other approval processing times to be short, for example none of the 'external' authorities or data owners required additional information or motivations.

Finland

Statistics Finland maintains population registration data, which include individual-level information on sociodemographic characteristics, housing, education, household and employment relationships as well as wages and salaries. Data on cause of death are retrieved from the death register held by Statistics Finland. Other health-related data are retrieved from two register holders: the National Institute for Health and Welfare provides data from the hospital discharge, social care, cancer and birth registers, whereas the Social Insurance Institution of Finland supplies data generated as part of the benefit administration process, including registers on medication reimbursement, sickness allowance and state pension.

Applications for research data are filed separately to each register-holding institution. All register holders make decisions independently regarding their own data. It is typically not necessary to obtain ethical permission prior to contacting the register holders; the decision process at each institution includes ethics vetting. Register holders of health data request statements from their own register experts and the Data Protection Ombudsman (similar to Data Protection Authorities in Norway). Once the applications have been approved, Statistics Finland delivers the PINs of the sampled individuals to the other register holders for data collection. The actual data linkage is then carried out by Statistics Finland. Most data sets are released to researchers in the remote-access platform of Statistics Finland. In contrast to hand-out data sets where all direct and indirect identifiers are removed, the remote-access platform allows access to data sets where only the direct identifiers are removed and the indirect identification of individuals is prevented by prohibiting outputs regarding small groups of people.

The data needs of the C-LIFE project partly overlapped with those of other ongoing research projects at the Population Research Unit at the University of Helsinki, and therefore two existing data sets could be used as the basis for research within C-LIFE. To comply with the data requirements of the C-LIFE project, the research team applied for these data sets to be updated as regards follow-up time and extended to cover new data sources. The data sets include (1) an 11% random sample of the population aged $\geqslant$15 years and permanently residing in Finland at the end of any of the years 1987–2007, supplemented with an 80% oversample of deaths in 1988–2007, and (2) a 20% random sample of Finnish households with children aged $<$15 years in 2000. The application regarding data set 2 also sought to extend the 20% sample.

Because the applications aimed to extend existing data sets to cover variables from new data sources and new research topics, updated research plans were needed. In the case of data set 1, the application sought to cover new variables from the registers of Statistics Finland, including information on residential areas, country of birth, more specific information on income, monetary transfers and social assistance, and crime. These data have recently become more available to researchers because of the increasingly utilised remote-access system. The requested updates and extensions were obtained with an approximate net processing time of 14 months. In the case of data set 2, where a sample extension was also applied for, the first step was to submit an updated research proposal to Statistics Finland. The updated data from Statistics Finland were obtained within three months from submission of the application. The data extraction and delivery was facilitated by the use of ready-made microdata modules recently introduced by Statistics Finland. After approval from Statistics Finland, applications for health data were filed to the other register holders. The net application processing time at the National Institute for Health and Welfare was one month, and the data were extracted and released in the remote-access system after an

additional six months. The application processing time at the Social Insurance Institution of Finland took three months, and the data were delivered four months later.

Updating and extending existing data sets that have been used in the C-LIFE project has been faster and more straightforward than obtaining a new data set. For example, a data update does not require statements from the Data Protection Ombudsman, which otherwise could extend the application processing time as much as six months. It is not uncommon that register holders also contact the research group during the data compilation process for further detail and negotiation. Furthermore, the application process within the C-LIFE project was facilitated by consultation with the research services department of Statistics Finland already before filing the formal application.

Norway

Statistics Norway holds data on demography, social conditions, social security, employment, rehabilitation and activation, education, income and wealth. As Statistics Norway is not allowed to store health register data, these need to be delivered from respective data owners. In our case, this included four agencies: the Medical Birth Registry and the Causes of Death Registry managed by the Norwegian Institute of Public Health, the Norwegian Patient Registry administrated by the Norwegian Directorate of Health and the Cancer Registry of Norway. Research projects need to submit separate applications to all register owners, as they are governed by separate legal frameworks. In addition, to access data on diagnoses associated with health-related benefits, a separate application had to be sent to the Norwegian Work and Welfare Administration.

Any social-science research project using non-anonymous data needs to submit a notification to the Norwegian Centre for Research Data (NSD). Based on the notification from the research institution, the NSD writes a DPIA in accordance with the GDPR. The research institution's data protection officer then approves the DPIA. Before submitting the data applications, dispensations from professional secrecy requirements must be approved by the Norwegian Work and Welfare Administration regarding diagnoses for health-related benefits, and from The Regional committee for medical and health research ethics (REC) regarding the use of health registers.

The Norwegian data application included the population resident in Norway from 1967 to 2016, as defined in the CPR. In total, about 100 health variables were included in the final application, and diagnosis included three digits in the ICD classification. In order to reduce the risk of indirect identification, data on any geographical information were omitted and country of birth was grouped into regions, as well as some reduction of detail in high income groups, type of education, occupation and birth date. Because the data owners have strict rules for handling each other's data, a decentralised data-merging procedure was chosen. Statistics Norway was to extract the population and assign a unique ID number to each individual in the population. Using this number and its link to the PIN, each data owner was to deliver data directly to the project. The project data do not include the PIN, as this would otherwise make direct identification possible. The project data manager would subsequently merge all the data using the ID number in a safe external storage environment.

The initial application was submitted before the GDPR had been implemented. At the time, the Data Protection Authority (Datatilsynet) was in charge of the ethical vetting, which is now done in the DPIA. As a result, experiences were gathered from two different application regimes. Net processing time for the initial data application was six months with the NSD and then seven months with Datatilsynet, ending with refusal. The main reasons for the refusal were that purposes were deemed too wide, processing time too long and the risk of identification too high. The REC had some similar concerns but invited us to provide clarifications. Revisions of the application included reduction of the number of purposes and the number of health variables, omission of the geographical variable, business region, shorter processing time and stronger emphasis on the expected value of the research. In the second round – still within the old regime – the corresponding figures for processing time were two months with the NSD and two months with Datatilsynet, concluding that no application submitted during the last months would receive a conclusion due to the implementation of GDPR during the spring of 2018. Hence, we needed to reformat our application to fit with the new application regime. In the third round, net processing time was six months with the NSD. The resulting DPIA could then be signed by the research institution, OsloMet, and attached with the data applications to the register owners. The regional committee for medical and health research ethics spent four months, including a follow-up inquiry to the project leader. The initial application to NAV was processed and approved within a few weeks. Statistics Norway spent four months before delivering the data. Data have also been delivered from the four health registers, with the Patient Register having the longest processing time – six months. Total net processing time for the Norwegian data, not counting parallel

Table I. Some aspects of the application process in DK, FI, NO and SE.

	Denmark	Finland[b]	Norway	Sweden
Net processing time (days)	100/150[a]	210	848/311[c]	435
Costs in 1000s €	6	25	41	10
Number of decision makers	5	3	7/6[d]	3/4[e]
Number of data retrievers	2	3	5	2

[a]The processing time was approximately 100 days from submission of first application and approximately 150 days including consultancy and processing time together with the Public Health Database staff.
[b]The figures refer to the process of updating and extending data set (2).
[c]The numbers refer to the total net processing time since the initial application and since the final application within the GDPR regime, respectively.
[d]The number of decision makers in old versus new regime
[e]Unlike the current regime, the ethical application was processed by both the regional and central ethical review boards.

processes, was 27 months. Counting from the start of the third round, however, the corresponding number is 10 months. In addition to these numbers, a considerable amount of time was spent by the project on writing and rewriting applications, filling out forms and follow-up in general.

In sum, the Norwegian data-acquisition process was very time-consuming and involved many decision makers. Seven different decision makers were involved, and eight in principle similar applications were submitted through separate forms. Some decision makers do comprehensive assessments of the project (NSD, SBB, REK and the Patient Register), while others seem to have more administrative processes. With the implementation of the GDPR, however, the process seems to have become more flexible and effective. Nevertheless, data acquisition is still cumbersome and costly (see Table I).

Sweden

Swedish register data are kept by different national agencies. Social data are kept by Statistics Sweden (*Statistiska Centralbyrån*; SCB), and health data are kept by the National Board of Health and Welfare (*Socialstyrelsen*; SoS). Besides these agencies, we acquired data on medically assisted reproduction treatments from the Q-IVF register. This is a national quality register, of which there are more than 100 in Sweden, which are specialised national registers monitored and certified by a national executive committee.

According to the law of ethical examination (*Lagen (2003:460) om etikprövning av forskning som avser människor*), research conducted on sensitive personal data on human subjects needs to pass ethical vetting. With effect from 1 January 2019, ethical vetting is conducted by a central agency, the Swedish Ethical Review Authority (*Etikprövningsmyndigheten*). The ethical vetting described here was carried out under the previous system using six regional review boards and one central review board, in cases where the regional review boards are unable to decide. Furthermore, in order to receive data from the Q-IVF register, a supplement to the ethical permission was obtained. This was done by writing a letter of motivation to the regional review board and paying a processing fee.

The process of acquiring linked register data on health and social conditions data in Sweden comprises several steps. Before the responsible register holders are contacted, ethical permission needs to be obtained. The regional review board of Stockholm deemed that the research project, focusing on the interplay between health and social conditions throughout the life course and across generations, could potentially be regarded as several separate projects. Consequently, they forwarded the case to the central review board. As the regional review board forwarded the case to the central review board, the project group supplemented the application with a letter. In the letter, it was argued that the described research is one, albeit broad, project and that carrying out the research as separate projects would limit the possibilities of comprehensive life-course analyses. The central ethical review board subsequently approved the application (Dnr Ö 25-2017). The process of obtaining ethical permission that enabled the applications for health and social data took just over two months once submitted, and the supplement needed to acquire data from the Q-IVF was approved within weeks of submitting.

Separate applications were made to the respective authorities that make independent assessments of confidentiality in accordance with theGDPR (replacing the law of personal information (*Personuppgiftslagen (1998:204)*) on 25 May 2018). SCB started processing the application around three months after the application was submitted. In some cases, the level of detail of variables was limited to reduce the risk of identification of individuals. These variables include country of birth and place of residence, which were delivered in grouped format. Most notably, SCB were reluctant to provide information on the day or month of birth. In the end an agreement was reached in which we were provided the year and month of birth for the majority of the population, but not the date. The data were delivered shortly after the confidentiality assessment was completed.

The evaluation of confidentiality at SoS is different compared to SCB, as health data, by definition, is considered to be sensitive personal information

under Swedish law and thereby surrounded by more rigorous legal protection. The evaluation was initiated seven months after the application was filed and is currently in its final stages, with a processing time of about five months.

In cases where all data are provided by one authority, data are often accessed through remote-access systems. However, in this case, the data are delivered by each authority separately to the facilities at the Department of Public Health Sciences, Stockholm University (SU). This is because the authorities do not deliver data to each other. The population is defined by SCB which generate a key for converting the PIN to an ID number that is unique to each individual. The key is shared by the different authorities but not with the researchers. The ID numbers are used to link the data once they have been delivered. The facilities at SU are approved for this purpose by the Swedish Data Protection Authority (*Datainspektionen*). The conditions are, however, that the data do not leave the facilities and are not accessed remotely. Similar agreements are signed with the register holders before data delivery.

Acquiring access to Swedish register data involves lengthy processes and interaction with multiple authorities and several different laws. Correspondence with the ethical review boards was the most straightforward and transparent. After each meeting where the current case was processed, signed written statements on the decision and motivation were received. Correspondence with the register holders, involving negotiations of data content, was more direct, informal and flexible, although less transparent, as clear motivations why restrictions and changes were deemed necessary were not always provided. Changes in both the legal framework (the introduction of the GDPR) and institutional arrangement (the establishment of the Swedish Ethical Review Authority) took place during the process of acquiring data. These changes illustrate the difficulty, from the point of view of researchers, to streamline the process of applying for data, since it requires navigating an ever-changing landscape of laws, agencies and administrative practices.

Comparing the processes in the different countries

The possibility of conducting comparative research rests on the condition that similar data can be accessed simultaneously in separate countries. This involves interacting with multiple stakeholders within different legal frameworks surrounding ethical permissions and the protection of personal integrity, as well as using different technical solutions. There are some similarities across countries in the procedures for obtaining access to combined health and socio-economic register data. All countries have strict ethical vetting procedures, including adherence to the need-to-know principle and limitations on data detail and data handling in order to prevent direct or indirect identification of the registered individuals. In all countries, the data owners perform independent vetting of the projects, in addition to an initial ethical vetting, either by the research institution (Denmark and, in part, Norway) or by designated ethical review boards (Sweden and Finland). No countries practise 'one-shop' solutions for accessing combined health and socio-economic resources; in all countries, several decision makers are involved.

In terms of processing time, costs, number of decision makers and data retrievers, the processes vary greatly. There also seem to be some variation regarding the ethical hurdles encountered with regard to the data needed for our project. As shown in Table I, processing time is generally very long in all countries, with the shortest time in Denmark (five months) and the longest in Norway (27 months) and Finland (30 months in the initial process preceding the C-LIFE project). However, we note that with the GDPR, a more effective regime seems to be in place in Norway, the last application round lasting about 10 months, which is comparable to the other countries. There are several decision makers involved in all countries. Again, Norway stands out with the most fragmented decision structure and thus with the largest administrative burden on the researchers. This is largely driven by the high number of data owners, with four different suppliers of health-related data in addition to Statistics Norway, all requiring separate applications. Costs vary drastically across countries, with Norway having costs almost seven times higher than Denmark. The lion's share of the Norwegian costs was demanded by Statistics Norway. Despite what seem like comparable pricing regimes in Sweden, Swedish costs are but a fraction of the Norwegian costs.

It is difficult to evaluate differences in ethical vetting across countries, but it seems as if within the old regime in Norway, a stricter understanding of general similar data protection rules [24] was practiced. The combination of sensitive data (health), family linkages and municipalities appear to be dismissed a priori (Datatilsynet, personal communication, 9 Februray 2016), and was not included in the initial application. This appeared to be less problematic in Denmark and Finland, possibly due to the availability of the institutionalised remote-access solution.

The introduction of the GDPR probably served to harmonise procedures to some extent, particularly in

Denmark and Norway, where in practice the university is responsible for and approves its own applications based on a DPIA. While Norway still has the most fragmented process in terms of the number of decision makers involved, the process has been greatly simplified by the changed roles of Datatilsynet. Instead of risking 'dead-end' applications from the NSD to Datatilsynet, as in the old regime, the project plan can now be edited in dialogue with the NSD and the Data Protection Officer at the university until the project is approvable. In Finland, there is a fresh initiative to establish a single-point application procedure.

Summing up, the Danish experience places itself at one end of a continuum with the fastest and cheapest process, while the Norwegian experience can be placed at the other end. In Nordic comparison, the Danish experience is recognised by excellent follow-up, quick processing, reasonable ethical vetting and low costs. However, the administrative burden on researchers is still an issue. The Norwegian situation was fragmented, work-intensive, strict, costly and very time-consuming, although there are strong signs of improvement. The Swedish and Finnish experiences also include long processing times.

Making data comparable

Register data are recorded and stored for administrative purposes, and are not primarily intended for research. Register data thus reflect the way policies and services are designed and how they operate. Political regulations influence what is registered and how. Comparing and contrasting Nordic countries allows new insights into the role of the welfare state in shaping inequalities, living conditions and health of populations. From this perspective, the differences between the countries in terms of how the respective institutions are organised and operate comprise a fundamental reason for undertaking comparative studies. However, these differences may also create comparability problems. For example, when comparing morbidity using register data on hospitalisation from different countries, the data will reflect both the level of morbidity in the population and differences in the health-care systems as well as differences in administrative practices, for example economic incentives that influence the registrations of hospitalisations and diagnoses, or patterns of entry and discharge [22]. Hence, even if there were no differences in the data-acquisition processes between countries and participating researchers, there are important steps to take in order to transform raw register data to proper research data. In short, data have to be made comparable. There are two main issues that are important to highlight in this context: differences in data driven by *differences in the social and welfare systems* that they reflect, and differences in data (that should be similar or identical) driven by differences in *procedures and practices* relating to data collection.

The first of these issues represents a general challenge to researchers using register data. Rather than being able to design and construct measures and indicators of the phenomena under study, researchers have to work with the information contained in the registers. As laws and regulations change, so do the data. For example, the abolishment of the wealth tax in Finland in 2006 and Sweden in 2007 ended the administrative need for information on wealth, and the wealth registers were therefore terminated. However, most changes in registrations are more subtle and involve, for example, reclassifying specific labour-market programmes into educational programmes or changes in the pension system or sickness insurance programmes. Changes in legislation, institutional design and the practices regarding both the policies per se as well as the collection of data influence the measurement of socio-economic position and computation of many measures of economic inequality, making it difficult to obtain consistent estimates within countries over time. In the context of comparative studies, this problem is amplified by differences between the countries. A key challenge is therefore to construct theoretically relevant variables that capture the phenomenon under study equally well in all countries.

This process often involves the construction of variables of interest by combining or transforming raw data in different ways. In some cases, this might involve quite simple transformations, such as adjusting income for purchasing power parities or classifying the educational systems into the ISCED standard [25]. In both these cases, the level of detail is reduced in order to ensure comparability. However, when the research questions require greater detail, for example involving a comparison of the economic value of social insurance systems in different countries, the construction of relevant research data is a research undertaking in itself. While this is a natural part of the research process regardless of the data used, it often becomes particularly complex in the case of administrative register data.

These are issues that may apply mainly to administrative data on socio-economic conditions, and there are register data that could be expected to be standardised and ready to use. However, in cases where the underlying phenomenon is the same, differences in the application of procedures for data collection and coding can vary a lot between countries

as well as within countries over time. An important example of this is cause of death statistics where several comparability problems can be pointed out, despite the fact that all Nordic countries follow international guidelines.

The cause-of-death registers contain information on underlying and contributing cause of death. All four countries apply the guidelines specified in the international classification of disease, currently in the 10th iteration. Besides providing a standardised classification of diseases, the guidelines also specify how to classify underlying and contributing cause of death. Both the quality of the registers [26] and how the data are collected differ between countries [27–30]. The most obvious difference is perhaps the difference in levels and trend of autopsy rates [31]. Although developments in medical technology lead to a declining need to perform autopsies to determine the cause of death, the possibility remains that differences in examination practices may lead to systematic differences in the composition of cause of death. For example, autopsy rates have been linked to suicide mortality [32]. Evaluations of errors on death certificates [33,34] indicate that a large proportion of errors are due to the underlying cause of death being reported as a contributing cause of death and vice versa, indicating that physicians may interpret the ICD guidelines differently. Finland has the highest autopsy rate out of the Nordic countries, and death certificates are systematically evaluated by a panel of nosologists [28] – no doubt factors that contribute to the Finnish cause-of-death register being regarded as one of the most accurate in the world [26]. Despite using the same standard for classification and reporting, there may then be systematic differences in quality and content of the Nordic cause-of-death registers.

In conclusion, access to data from two or more countries is necessary but not sufficient to do comparative analyses. Even data that should be directly comparable must be checked and corrected in order to make valid comparisons. In our case, with research questions that involve the interplay between social, economic and health conditions, there is a need to construct theoretically driven variables that reflect the constructs that we are interested in, for example the degree to which welfare-state programmes cover income losses due to illness. This, in turn, means that data in many cases need to be transformed, combined and grouped in a way that makes them less sensitive than the raw data. By using similar guidelines in each country, this process also provides a possibility to promote comparability. On the other hand, sufficient detail in the raw data is necessary to allow for flexibility in constructing the comparative classifications.

Modes of collaboration using Nordic social and health register data

In this section, we review a number of approaches to work with data in comparative projects. We have identified six different ways to perform pooled and comparative analyses of Nordic register data, but we have not been able to use all six in the project.

Local coordinated analyses. The article 'Income security in Nordic welfare states for men and women who died when aged 55–69 years old' [35] made use of this method. This paper relied on individual-level data that identified all deaths in selected years by age and sex. For these individuals, annual information on social characteristics in the years before death was extracted, in particular income and household composition. Analyses were carried out locally in each country with analyses codes provided by the first author of the paper. This mode of collaboration provides flexibility in developing and applying statistical models for individual data, but it is demanding, as time and expertise need to be available at all participating centres and the comparison of country-specific estimates can be more challenging. Participant confidentiality is guaranteed by local data-access regulations.

Central analyses based on tabulated data. The paper 'Contribution of smoking and alcohol consumption to income differences in life expectancy: evidence using Danish, Finnish, Norwegian and Swedish register data' by Östergren et al. [36] utilised the tabulated data approach. The data consisted of aggregate level tables with deaths and person-years cross-tabulated according to the variables of interest – in this particular case, five-year age group, sex, calendar period and income quintile. Such tables were constructed for each country and analysed centrally. Participant confidentiality is guaranteed by the aggregate-level nature of the data and restrictions on small cell counts. The approach is feasible for tightly formulated research questions, for example on the trends and changes in social inequalities of health between countries, but it does not allow for analyses of individual-level dynamics of changes in social circumstances and health.

Remote access. The remote-access solution is put in place in several institutions/authorities in the Nordic countries. It is, in principle, possible to attach and authorise researchers affiliated with what in the single country is regarded as a safe institution (e.g. a researcher in an institution authorised to work on data in Statistics Finland) to a sister institution in another Nordic country (e.g. Statistics Denmark).

The advantage is that a project can be driven forward by the researcher who has the idea and the skills to make the demanded analyses, to develop and change analytical strategies according to learnings from the process and to secure comparability. Disadvantages may include costly access fees, cumbersome import of external data or software, delays on output due to monitoring and difficulties in getting access for collaborators. This solution does not, however, entail an exchange or pooling of the data and thus resemble the 'local coordinated analyses' approach but with the benefits of having one person responsible for conducting the analyses (in cooperation with the local researchers).

Transfer of data between safe environments. The Nordic National Statistical Institutions (NSIs) have indicated that a common model of cooperation has been made between the statistical institutions enabling social data gathered by the statistical institutions to be transferred for specific purposes, when the need for pooled data is justified (see www.nordman.network). This would be the case when the phenomenon studied is important but rare, for instance, when studying rare exposures (e.g. having a very old father) or an infrequent but serious outcome (e.g. childhood cancer or homicide). This approach is only valid when the result most likely applies to all populations that are pooled in the study. To date, the collaboration only applies to data gathered by NSIs (social data) and thus does not cover health data, which is an important drawback. Moreover, the cooperation model requires that access to data is given only through a remote-access solution, which is at the moment not used in Norway.

Large Nordic harmonised pooled data set. In principle, data from the national health registers and NSIs could be exchanged/pooled and harmonised. This has been done through well-established collaboration both for medical birth registers [37,38], for cancer registers (NordCan) [39] and for social data (ALiCCS) [40]. The idea of creating a large Nordic harmonised data set is appealing. Pooling data makes possible statistical analyses of rare diseases (e.g. childhood cancer) and small groups (e.g. immigrant cohorts from specific countries), and allows for more sophisticated control groups in individual-level analyses (e.g. matching procedures, controlling for country fixed effects). However, current data regulation practices make such a solution very difficult.

Co-analyses without data storage outside the local institutions. Various technical solutions have been developed to meet the need of analysing data stored securely at different places without physically exchanging or sharing the data. Common to these solutions is that data are stored locally, but are co-analysed using an analysis server that runs the functions on the individual-level data, but only returns outputs that are non-disclosive (e.g. summary statistics, score vectors and information matrices) – and only after approval from data owners and ethical authorities. An example of this system is DataShield [41]. Such methods of collaboration and data sharing are currently explored in EU-funded projects such as LifeCycle (https://lifecycle-project.eu/), ReCap (https://www.recap-h2020.eu/) and the Canadian-EU based project EUCAN-connect. In the future, experiences from these large-scale studies may be used to guide attempts to apply these or similar systems to Nordic register data.

Conclusions

Nordic register data are no doubt a potential gold mine for research. However, in our experience, a number of hurdles and bottlenecks still impede the anticipated 'gold rush'. The data-acquisition processes we document here show that long processing times, fragmented decision structures, a high administrative burden on researchers and ever-changing administrative frameworks pose serious threats to the realisation of comparative Nordic research projects. The fastest and cheapest process was experienced in Denmark, while Norway occupied the less favourable end of the continuum. The considerable cost of retrieving data from Statistics Norway may in itself represent a barrier to unfunded research collaborations. A limitation of this paper, however, is that pre-existing arrangements in Denmark and Finland, and a troublesome implementation of the GDPR in Norway, may affect the generalisability of our experiences. On the other hand, the obstacles identified were quite similar across countries, and are likely to apply to any similar new comparative project.

The favourable conditions found in Denmark may in part be due to the fact that health and social data can be stored at one single institution (Statistics Denmark). This is not possible in Norway and Sweden, although Finland's remote-access system allows this. Another contributing factor is probably the opportunity to establish a predefined data set (The Public Health Database) from which project data can be retrieved more swiftly. Third, the lack of additional external ethical vetting in Denmark, as practised in Sweden and Finland (and in the previous Norwegian application regime), is also likely to speed up the process. Finally, follow-up and guidance during the application work, experienced also in Norway, seems to be productive.

We have shown that comparability issue may be due to either differences in procedures and practices relating to data collection or differences in the systems. As registers are kept for administrative purposes, changes and differences in the institutional environment and political decisions impact what information is collected and stored. Specific practices surrounding data collection, reporting and validation influence the quality of the data. Third, what is included in the data may differ between the countries. One type of difference relates to the classifications used. System differences are less straightforward and relate to how information is affected by the overall institutional and social environment in each country. Both types of issues may be addressed by aggregating or reclassifying variables so that they indicate the same phenomena. Finally, the data-acquisition process may in itself limit the data made available to research projects and hence the comparability as a result of negotiations and imposed restrictions.

There are clear methodological benefits from pooling Nordic register data. However, typically register owners explicitly prohibit storage of data outside national borders. Although there are known cases (see above) in which this has been allowed, the slow progress of our project prevented us from exploring this opportunity further. Existing efforts, such as the NordMAN collaboration, did not meet the needs of this project, as it does not include the possibility of including health data. Yet, other options exist, as we show above. Local coordinated analyses, centralised analyses of tabulate data, remote-access solutions and new technical solutions for data sharing provides excellent opportunities for most purposes.

Our experiences have some implications for the organisation of research. In addition to all the work done with coordinating, writing and following up the applications, the processing time of more than two years (and still counting) to obtain all the data is incompatible with typical three-year projects favoured by many research funding bodies. Furthermore, data access can only be granted for clear and limited research purposes, so establishing data before being funded by a topical research call may be a challenge. In addition, some data owners and ethical boards are reluctant to approve more durable data infrastructure, which mean that manuscripts submitted towards the end of the funded project period may run into difficulties in responding to peer review.

Our experiences are not unique. A position paper[42] on the implementation of the GDPR from a Nordic network led by NSD sums up some of the challenges based on meetings with researchers (NSD 2017). In particular, the paper focussed on the unnecessary long and repetitive application processes (p. 5). In line with the intentions of the GDPR, the paper argues that Nordic Ministers should make it explicit that 'one data protection impact assessment from one Nordic country is sufficient for data owners to grant access to cross-border processing of personal data for scientific research purposes' (p. 6). In line with previous calls [16,42], the paper also proposes that a clear legal basis for pooling Nordic register data should be provided. We strongly support these suggestions, which would make comparative research much easier. However, we note that few signs of Nordic harmonisation could be observed in the first year of the GDPR. Thus, the 'gold mine' for research has yet to see its Klondike, and remains a harsh place, even for the most dogged of miners.

Acknowledgements

The experiences we draw on in this paper are the result of work done by the entire C-LIFE project. In addition to the authors listed, we would like to acknowledge the contribution to the project made by Abdulfatah Adam, Kersti Bergqvist, Josephine Funck Bilsteen, Henrik Brønnum-Hansen, Magne Bråthen, Espen Dahl, Jon Ivar Elstad, Karl Gauffin, Anne Vinkel Hansen, Åsmund Hermansen, Taina Leinonen, Heta Moustgaard, Lasse Tarkiainen and David Taylor-Robinson.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This work was funded by NordForsk (project number 75970).

ORCID iDs

Kjetil A. Van Der Wel https://orcid.org/0000-0002-1812-4795
Olof Östergren https://orcid.org/0000-0002-7156-3260

References

[1] Bartley M. *Health inequality: an introduction to theories, concepts, and methods*. Cambridge: Polity Press, 2004.
[2] Marmot M. *The health gap. The challenges of an unequal world*. London: Bloomsbury, 2015.
[3] Dahl E, Bergsli H and van der Wel K. *Sosial ulikhet i helse: en norsk kunnskapsoversikt*. Oslo: Oslo and Akershus University College, 2014.
[4] Diderichsen F, Andersen I, Manuel C, et al. Health inequality – determinants and policies. *Scand J Public Health* 2012;40:12–105.

[5] Lundberg O. *Nästa steg på vägen mot en mer jämlik hälsa – Förslag för ett långsiktigt arbete för en god och jämlik hälsa*. Stockholm: SOU, 2017.
[6] Kuh D, Ben-Shlomo Y, Lynch J, et al. Life course epidemiology. *J Epidemiol Community Health* 2003;57:778–783.
[7] Ferrarini T, Nelson K and Sjoberg O. Decomposing the effect of social policies on population health and inequalities: an empirical example of unemployment benefits. *Scand J Public Health* 2014;42:635–642.
[8] Mackenbach J and de Jong J. *Health inequalities. An interdisciplinary exploration of socioeconomic position, health and causality*. Amsterdam: The Royal Netherlands Academy of Arts and Sciences, 2018.
[9] Fritzell J and Lundberg O. *Health inequalities and welfare resources: continuity and change in Sweden*. Bristol: Policy Press, 2007.
[10] Hillier-Brown FC, Thomson K, McGowan V, et al. The effects of social protection policies on health inequalities in European welfare states: a review of systematic reviews. *Scand J Public Health*. Epub ahead of print 9 May 2019. DOI: 10.1177/1403494819848276.
[11] Thomson K, Hillier-Brown F, Todd A, et al. The effects of public health policies on health inequalities in high-income countries: an umbrella review. *BMC Public Health* 2018;18:869.
[12] Bambra C. Health inequalities and welfare state regimes: theoretical insights on a public health 'puzzle'. *J Epidemiol Commun Health* 2011;65:740–745.
[13] Mackenbach JP. Nordic paradox, Southern miracle, Eastern disaster: persistence of inequalities in mortality in Europe. *Eur J Public Health* 2017;27:14–17.
[14] Pförtner TK, Pfaff H and Elgar FJ. The role of welfare state characteristics for health and inequalities in health from a cross-national perspective: a critical research synthesis. *Koln Zeitschr Soziol Sozialpsychol* 2019;2019:1–25.
[15] Brennenstuhl S, Quesnel-Vallee A and McDonough P. Welfare regimes, population health and health inequalities: a research synthesis. *J Epidemiol Community Health* 2012;66:397–409.
[16] Könberg B. *Det framtida nordiska hälsosamarbetet*. Reykjavik: Nordic Council, 2014.
[17] Langhoff-Roos J, Krebs L, Klungsoyr K, et al. The Nordic medical birth registers – a potential goldmine for clinical research. *Acta Obstet Gynecol Scand* 2014;93:132–137.
[18] Nordforsk. *Nordic biobanks and registers. A basis for innovative research on health and welfare*. Policy paper. Oslo: Nordforsk, 2017.
[19] Thygesen LC, Daasnes C, Thaulow I, et al. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand J Public Health* 2011;39:12–16.
[20] United Nations. *United National Economic Commission for Europe. Register-based statistics in the Nordic countries*. New York: United Nations, 2007.
[21] Erlangsen A and Fedyszyn I. Danish nationwide registers for public health and health-related research. *Scand J Public Health* 2015;43:333–339.
[22] Maret-Ouda J, Tao W, Wahlin K, et al. Nordic registry-based cohort studies: possibilities and pitfalls when combining Nordic registry data. *Scand J Public Health* 2017;45:14–19.
[23] Nordforsk. *Nordic eScience Action Plan 2.0. eScience and eInfrastructure in an international context*. Policy paper. Oslo: Nordforsk, 2015.
[24] Salokannel M. *Ethical review, data protection and biomedical rsearch in the Nordic countries: a legal perspective*. Policy paper. Oslo: Nordforsk, 2017.
[25] Schneider SL (ed). *The International Standard Classification of Education (ISCED-97): an evaluation of content and criterion validity for 15 European countries*. Mannheim, Germany: MZES, 2008.
[26] Phillips DE, Lozano R, Naghavi M, et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Popul Health Metr* 2014;12:14.
[27] Helweg-Larsen K. The Danish register of causes of death. *Scand J Public Health* 2011;39:26–29.
[28] Lahti RA and Penttilä A. The validity of death certificates: routine validation of death certification and its effects on mortality statistics. *Forensic Sci Int* 2001;115:15–32.
[29] Pedersen A and Ellingsen C. Data quality in the causes of death registry. *Tidsskr Nor Laegeforen* 2015;135:768–770.
[30] Brooke HL, Talbäck M, Hörnblad J, et al. The Swedish cause of death register. *Eur J Epidemiol* 2017;32:765–773.
[31] Tøllefsen IM, Hem E, Ekeberg Ø, et al. Differing procedures for recording mortality statistics in Scandinavia. *Crisis* 2017;38:123–130.
[32] Kapusta ND, Tran US, Rockett IR, et al. Declining autopsy rates and suicide misclassification: a cross-national analysis of 35 countries. *Arch Gen Psychiatry* 2011;68:1050–1057.
[33] Mieno MN, Tanaka N, Arai T, et al. Accuracy of death certificates and assessment of factors for misclassification of underlying cause of death. *J Epidemiol* 2016;26:191–198.
[34] Cheng T-J, Lin C-Y, Lu T-H, et al. Reporting of incorrect cause-of-death causal sequence on death certificates in the USA: using hypertension and diabetes as an educational illustration. *Postgrad Med J* 2012;88:690–693.
[35] Elstad JI, Hermansen Å, Brønnum-Hansen H, et al. Income security in Nordic welfare states for men and women who died when aged 55–69 years old. *J Int Compar Soc Pol* 2019;35:157–176.
[36] Ostergren O, Martikainen P, Tarkiainen L, et al. Contribution of smoking and alcohol consumption to income differences in life expectancy: evidence using Danish, Finnish, Norwegian and Swedish register data. *J Epidemiol Commun Health* 2019;73:334–339.
[37] Mortensen LH, Diderichsen F, Arntzen A, et al. Social inequality in fetal growth: a comparative study of Denmark, Finland, Norway and Sweden in the period 1981–2000. *J Epidemiol Commun Health* 2008;62:325–331.
[38] Petersen CB, Mortensen LH, Morgen CS, et al. Socio-economic inequality in preterm birth: a comparative study of the Nordic countries from 1981 to 2000. *Paediatr Perinat Epidemiol* 2009;23:66–75.
[39] Engholm G, Ferlay J, Christensen N, et al. NORDCAN – a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol* 2010;49:725–736.
[40] Asdahl PH, Winther JF, Bonnesen TG, et al. The Adult Life after Childhood Cancer in Scandinavia (ALiCCS) study: design and characteristics. *Pediatr Blood Cancer* 2015;62:2204–2210.
[41] Wilson R, Butters O, Avraam D, et al. DataSHIELD – new directions and dimensions. *Data Sci J* 2007;16:1–21.
[42] Kvalheim V and Myhren MH. New legislation – a unique opportunity for harmonising the legal framework for research in the Nordic countries. Bergen: NSD – Norwegian Centre for Research Data/Nordic Network of Researchers and Research Infrastructure Experts, 2017.